The Battle of Neighbourhoods: Safety, Relocators and Immigrants

By:
Prashant Pillai

INTRODUCTION

Background

I recently came across an article that was based on a survey where various relocation service providers (more commonly knows as packers and movers) were questioned about their operations and various trends that they usually observe when their customers relocate from one place to another. The primary objective was to study the customer's mindsets and figure out what triggered the initial move and then, what was the most prominent factor for them while choosing the location that they were moving to. When it came down to it, it was observed that it was a toss-up between price and the safety of the neighbourhood.

Problem

The problem was based in London, that was triggered by observing a sharp uptake in the number of immigrants and relocators in the recent years. Further probing resulted in a trend where certain neighbourhoods seemed to attract a major chunk of these relocators and immigrants. Thus, I thought it would be interesting to analyse London's neighbourhoods based on how safe they are and see if there was a correlation between it and the findings of the survey.

Stakeholders/Audience

If there indeed is a correlation between the survey and our findings, it can serve as a template or point of reference for other people looking to relocate or new immigrants and help them find a neighbourhood to their liking. Also, since living spaces are limited, there might be some safe neighbourhoods that might be full and cannot accommodate any more people. On the other hand, there might be some neighbourhoods that are equally safe but haven't been occupied yet because of their low-key status. In such scenarios, it can help the audience find their second or third best alternatives without too many compromises.

DATA

Data Acquisition

The data acquired for this project is a combination of data from three sources. The first data source of the project uses a London crime data that shows the crime per Borough in London. The dataset contains the following columns:

- lsoa_code: code for Lower Super Output Area in Greater London.
- Borough: Common name for London Borough.
- major_category: High level categorization of crime
- minor_category: Low level categorization of crime within major category.
- value: monthly reported count of categorical crime in given Borough
- year: Year of reported counts, 2008-2016
- month: Month of reported counts.

The second source of data is scraped from a Wikipedia page that contains the list of London Boroughs. This page contains additional information about the Boroughs with the following columns:

- Borough: The names of the 33 London Boroughs.
- Inner: Categorizing the Borough as an Inner London Borough or an Outer London Borough.
- Status: Categorizing the Borough as Royal, City or another Borough.
- Local authority: The local authority assigned to the Borough.
- Political control: The political party that control the Borough.
- Headquarters: Headquarters of the Boroughs.
- Area (sq mi): Area of the Borough in square miles.
- Population: The population in the Borough recorded during the year 2013.
- Co-ordinates: The latitude and longitude of the Boroughs.
- Nr. in map: The number assigned to each Borough to represent visually on a map.

The third data source is the list of Neighborhoods in the royal Borough of “Kingston upon Thames” as found on a Wikipedia page. This dataset is created from scratch using the list of neighborhoods available on the site with the following are columns:

- Neighborhood: Name of the neighborhood in the Borough.
- Borough: Name of the Borough.
- Latitude: Latitude of the Borough.
- Longitude: Longitude of the Borough.

Data Processing

The data preparation for each of the three sources of data is done separately.
From the London crime data, the crimes during the most recent year (2016) are selected. The major categories of crime are pivoted to get the total crimes per Borough, per the category.

The second data is scraped from a Wikipedia page using the Beautiful Soup library in python. Using this library, we extract the data in the tabular format that is displayed on the website. After the web scraping, string manipulation is used to get the names of the Boroughs in the correct form. This is important because we will be merging the two datasets together on the column name "Borough".

The two datasets are merged on the Borough names to form a new dataset that combines the necessary information in one dataset. The purpose of this dataset is to help in visualizing the crime rates in each Borough and identify the Borough with the least crimes recorded during the year 2016.

After visualizing the crime in each Borough, we can find the Borough with the lowest crime rate and hence tag that Borough as the safest Borough.

The third source of data is acquired from the list of neighborhoods in the safest Borough on Wikipedia. This dataset is created from scratch, the pandas data frame is created with the names of the neighborhoods and the name of the Borough. The 'latitude' and 'longitude' columns are left blank.

The coordinates of the neighborhoods are obtained using Google Maps Geocoding API to get the final dataset. The new dataset is used to generate the venues for each neighborhood using the Foursquare API.

	Borough	Local authority	Political control	Headquarters	Area (sq mi)	Population (2013 est)[1]	Co-ordinates	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
0	Barking and Dagenham	Barking and Dagenham London Borough Council	Labour	Town Hall, 1 Town Square	13.93	194352	51°33′39″N 0°09′21″E / 51.5607°N 0.1557°E	1287	1949	919	378	534	5607	6067	16741
1	Barnet	Barnet London Borough Council	Conservative	Barnet House, 2 Bristol Avenue, Colindale	33.49	369088	51°37′31″N 0°09′06″W / 51.6252°N 0.1517°W	3402	2183	906	499	464	9731	7499	24684
2	Bexley	Bexley London Borough Council	Conservative	Civic Offices, 2 Watling Street	23.38	236687	51°27′18″N 0°09′02″E / 51.4549°N 0.1505°E	1123	1673	646	294	209	4392	4503	12840
3	Brent	Brent London Borough Council	Labour	Brent Civic Centre, Engineers Way	16.70	317264	51°33′32″N 0°16′54″W / 51.5588°N 0.2817°W	2631	2280	2096	536	919	9026	9205	26693
4	Bromley	Bromley London Borough Council	Conservative	Civic Centre, Stockwell Close	57.97	317899	51°24′14″N 0°01′11″E / 51.4039°N 0.0198°E	2214	2202	728	417	369	7584	6650	20164

Figure representing the final data frame created from all three sources.

METHODOLOGY

Exploratory Data Analysis

Statistical summary of crimes

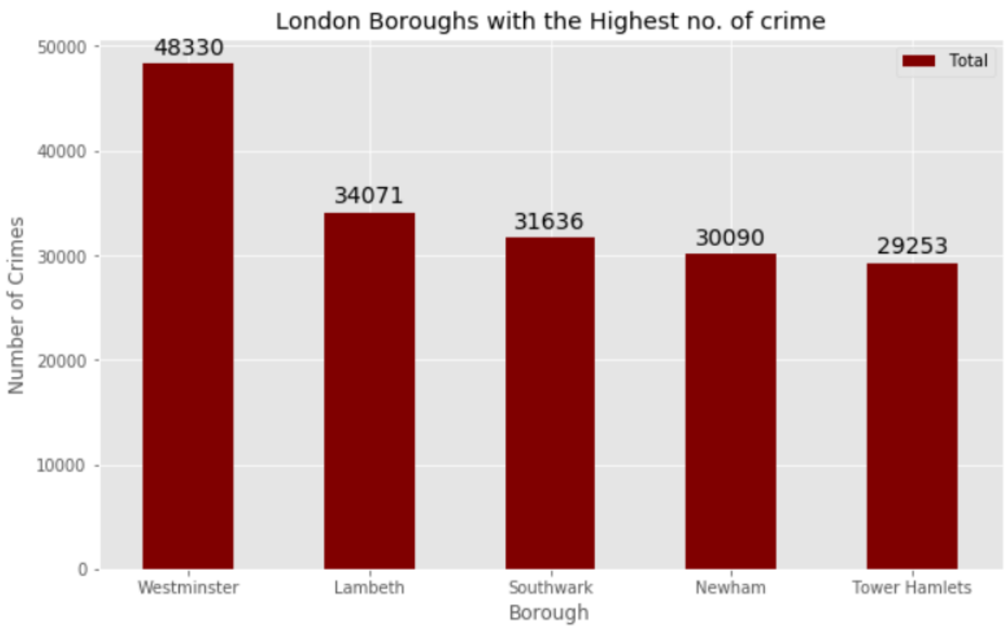
I used the .describe() function in Python to get statistics of the London crime data. The function returns the mean, standard deviation, minimum, maximum, 1st quartile (25%), 2nd quartile (50%), and the 3rd quartile (75%) for each of the major categories of crime. The count for each of the major categories of crime returns the value '33' which is the number of London boroughs. 'Theft and Handling' is the highest reported crime during the year 2016 followed by 'Violence against the person', 'Criminal damage'. The lowest recorded crimes are 'Drugs', 'Robbery' and 'Other Notifiable offenses.'

	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
count	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000
mean	2069.242424	1941.545455	1179.212121	479.060606	682.666667	8913.121212	7041.848485	22306.696970
std	737.448644	625.207070	586.406416	223.298698	441.425366	4620.565054	2513.601551	8828.228749
min	2.000000	2.000000	10.000000	6.000000	4.000000	129.000000	25.000000	178.000000
25%	1531.000000	1650.000000	743.000000	378.000000	377.000000	5919.000000	5936.000000	16903.000000
50%	2071.000000	1989.000000	1063.000000	490.000000	599.000000	8925.000000	7409.000000	22730.000000
75%	2631.000000	2351.000000	1617.000000	551.000000	936.000000	10789.000000	8832.000000	27174.000000
max	3402.000000	3219.000000	2738.000000	1305.000000	1822.000000	27520.000000	10834.000000	48330.000000

Figure representing the statistical summary of the crimes committed in Boroughs of London.

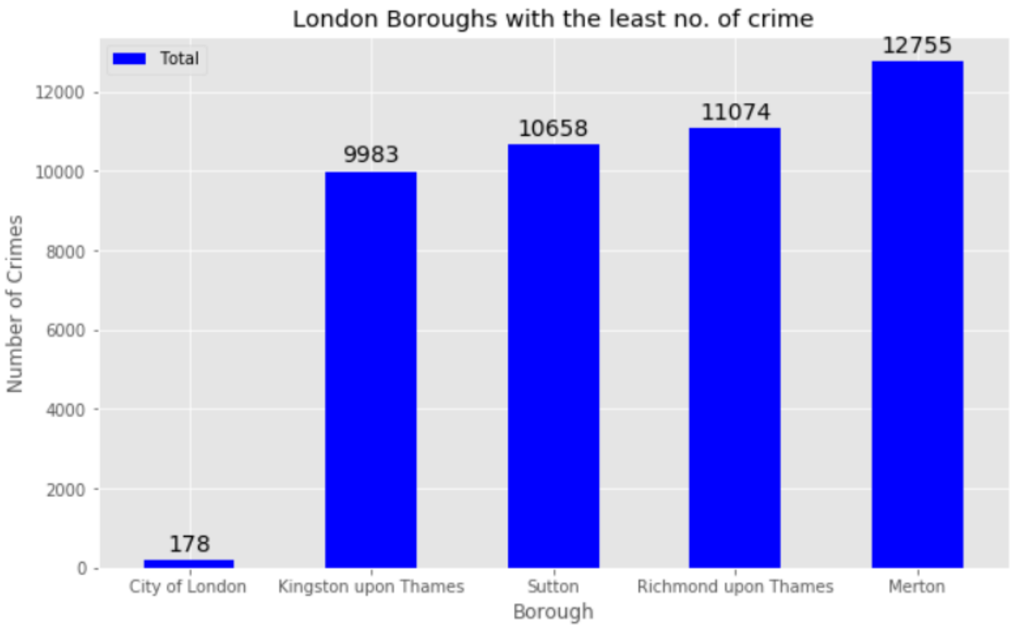
Boroughs with the highest crime rates

Comparing five boroughs with the highest crime rate during the year 2016 it is evident that Westminster has the highest crimes recorded followed by Lambeth, Southwark, Newham and Tower Hamlets. Westminster has a significantly higher crime rate than the other 4 boroughs.



Boroughs with the lowest crime rates

Comparing five boroughs with the lowest crime rate during the year 2016, City of London has the lowest recorded crimes followed by Kingston upon Thames, Sutton, Richmond upon Thames and Merton. City of London has a significantly lower crime rate because it is the 33rd principal division of Greater London but it is not a London borough. It has an area of 1.12 square miles and a population of 7000 as of 2013 which suggests that it is a small area. Hence, we will consider the next borough with the lowest crime rate as the safest borough in London which is Kingston upon Thames.



Neighborhoods in Kingston upon Thames

There are 15 neighborhoods in the royal borough of Kingston upon Thames, they are visualized on a map using folium on python.

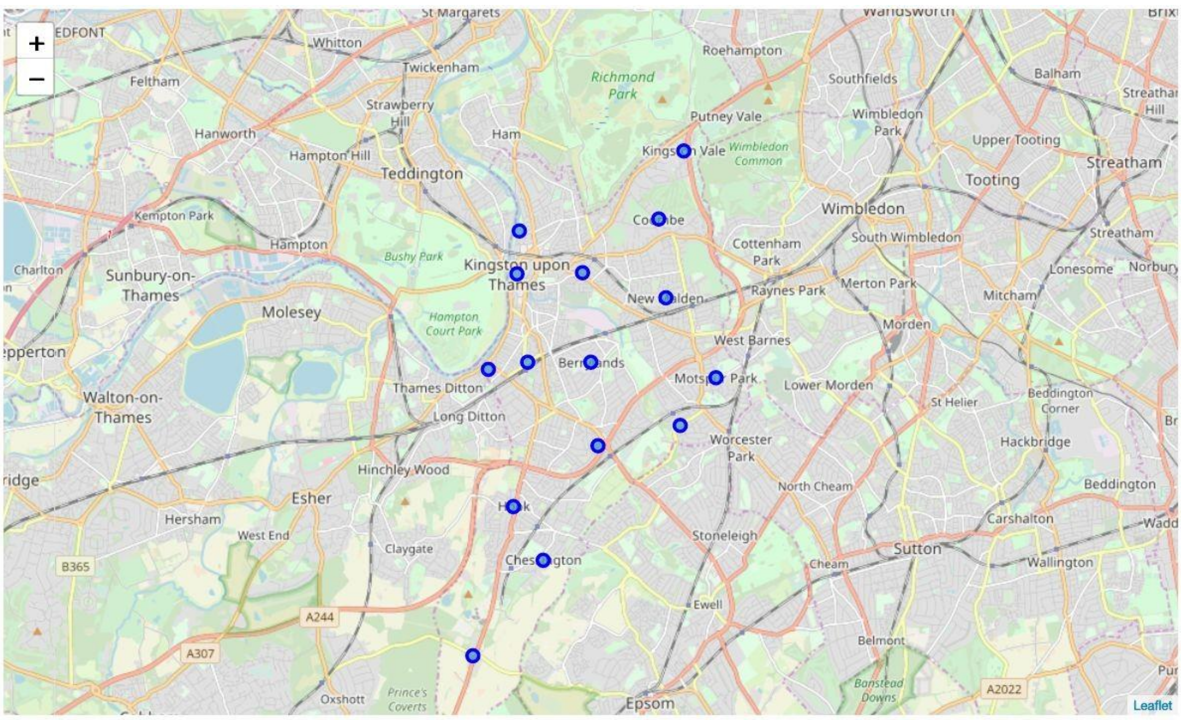


Figure representing the Neighborhoods in Kingston upon Thames.

Modelling

Using the final dataset containing the neighborhoods in Kingston upon Thames along with the latitude and longitude, I found all the venues within a 500-meter radius of each neighborhood by connecting to the Foursquare API. This returns a json file containing all the venues in each neighborhood which is then converted to a pandas data frame. This data frame contains all the venues along with their coordinates and category.

One hot encoding (a process by which categorical variables are converted into a form that could be provided to ML algorithms to do a better job in prediction) is done on the venues data. The Venues data is then grouped by the Neighborhood and the mean of the venues is calculated. Finally, the 10 common venues are calculated for each of the neighborhoods.

To help people find similar neighborhoods in the safest borough I cluster similar neighborhoods using K - means clustering, which is a form of unsupervised machine learning algorithm that clusters data based on predefined cluster size. I will be using a cluster size of 5 for this project that will cluster the 15 neighborhoods into 5 clusters. The reason behind conducting a K- means clustering is to cluster neighborhoods with similar venues together so that people can shortlist the area of their interests based on the venues/amenities around each neighborhood.

RESULTS

After running K-Means clustering, we can examine each cluster.

Upon further observation, we can see that clusters 1 and 4 have 2 neighbourhoods each, while clusters 3 and 5 have a single neighbourhood. Cluster 2 is the cluster that has the highest neighbourhoods and comprises of 7 of the total 13 neighbourhoods. The screenshots of each cluster are presented below:

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
6	Kingston Vale	Kingston upon Thames	51.431850	-0.258138	0	Grocery Store	Sandwich Place	Bar	Soccer Field	Deli / Bodega	Department Store	Discount Store	Dry Cleaner	Electronics Store	Farmers Market
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	0	Grocery Store	Pharmacy	Sandwich Place	Train Station	Hotel	Discount Store	Coffee Shop	Café	Bus Stop	Restaurant

Cluster-1

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Canbury	Kingston upon Thames	51.417499	-0.305553	1	Pub	Park	Fish & Chips Shop	Supermarket	Spa	Shop & Service	Gym / Fitness Center	Plaza	Indian Restaurant	Hotel
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	1	Coffee Shop	Café	Sushi Restaurant	Burger Joint	Pub	Asian Restaurant	Portuguese Restaurant	French Restaurant	German Restaurant	Electronics Store
9	New Malden	Kingston upon Thames	51.405335	-0.263407	1	Gastropub	Bar	Sushi Restaurant	Supermarket	Gym	Chinese Restaurant	Korean Restaurant	Indian Restaurant	Wine Shop	Farmers Market
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	1	Italian Restaurant	Food	Platform	Pub	Indian Restaurant	Breakfast Spot	Japanese Restaurant	Hotel	Hardware Store	Pharmacy
11	Old Malden	Kingston upon Thames	51.382484	-0.259090	1	Train Station	Food	Pub	Child Care Service	Construction & Landscaping	Fried Chicken Joint	French Restaurant	Furniture / Home Store	Fish & Chips Shop	Cosmetics Shop
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	1	Indian Restaurant	Café	Coffee Shop	Pub	Hotel	Chinese Restaurant	Italian Restaurant	Restaurant	Pet Café	Harbor / Marina
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	1	Coffee Shop	Pub	Grocery Store	Italian Restaurant	Pharmacy	Bakery	Tea Room	Gastropub	Train Station	Farmers Market

Cluster-2

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
4	Hook	Kingston upon Thames	51.367898	-0.307145	2	Indian Restaurant	Fish & Chips Shop	Bakery	Supermarket	Department Store	Discount Store	Dry Cleaner	Electronics Store	Farmers Market	Fast Food Restaurant

Cluster-3

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	3	Bus Stop	Gym / Fitness Center	Park	Wine Shop	Department Store	Discount Store	Dry Cleaner	Electronics Store	Farmers Market	Fast Food Restaurant
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	3	Bus Stop	Park	Gym	Soccer Field	Restaurant	Wine Shop	Department Store	Discount Store	Dry Cleaner	Electronics Store

Cluster-4

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	4	Grocery Store	Garden Center	Pub	Restaurant	Electronics Store	Deli / Bodega	Department Store	Discount Store	Dry Cleaner	Farmers Market

Cluster-5

DISCUSSION

The aim of this project was to help potential relocators and new immigrants in finalising a neighbourhood to move into, based on their lifestyles and preferences. For example, if there is someone who is into fitness and prefers a neighbourhood where it’s easier to get around without the need of buying a vehicle, Cluster-3 has two neighbourhoods that have gyms and parks as common venues along with bus stops.

CONCLUSION

When compared to the survey I found their results to be pretty close to my own findings with Sutton and Kingston upon Thames seeming to attract most of the people moving into these neighbourhoods. Thus, this project can encourage potential movers to go through a full research of the Boroughs and the various neighbourhoods within them. It would also help them get a basic understanding of what each neighbourhood has to offer in terms of the venues around the place and make an informed decision.